



OGS

National Institute
of Oceanography
and Applied
Geophysics

Scientific Council Meeting 2023

Open science for societal challenges:
the importance of marine data curation

Relatore
Marina Lipizer

Contributors: A. Altenburger, C. Altobelli, A. Cocciandich, M.A. French, A. Giorgetti, M.E. Molina Jack, M. Jokic, E. Kubin,
E. Partescano, N.C. Reyes Suarez, M. Vinci

Research and innovation

Home > Strategy on research and innovation > Strategy 2020-2024 > Our digital future > Open Science

Open Science

An approach to the scientific process that focuses on spreading knowledge as soon as it is available using digital and collaborative technology. Expert groups, publications, news and events.



Research and innovation strategy 2020-2024

Helping deliver the Commission's 6 goals

1. Environment and climate
2. Our digital future
3. Jobs and economy
4. Protecting our citizens and our values
5. Europe in the world
6. Democracy and rights

The EU's open science policy

Open science is a **policy priority** for the European Commission.

...When researchers **share** knowledge and data as early as possible in the research process with all relevant actors it helps diffuse the latest knowledge.

And when partners from across academia, industry, public authorities and citizen groups are invited to participate in the research and innovation process, creativity and **trust in science** increases.

... the Commission requires beneficiaries of research and innovation funding to make their **publications** available in **open access** and make their **data as open as possible and as closed as necessary**.



Home | Stampa | Notizie e comunicati stampa | Pubblicato il Piano nazionale della scienza aperta

Pubblicato il Piano nazionale della scienza aperta

Lunedì, 20/06/2022

Individuati 5 assi di intervento: pubblicazioni scientifiche, dati, valutazione della ricerca, partecipazione e apertura dei dati della ricerca su SARS-COV-2 e Covid-19

... fine data curation

Open Science in practice

...Open Science



Making science more **accessible**, **inclusive** and **equitable** for the benefit of all.

... a critical accelerator for the achievement of the United Nations **Sustainable Development Goals** and a true game changer in bridging the science, technology and innovation gaps and fulfilling the human right to science.

Involves OPEN:

- Publications
- Data
- Softwares
- Tools
- Educational resources
- Notebooks
- ...
- ...
- ...



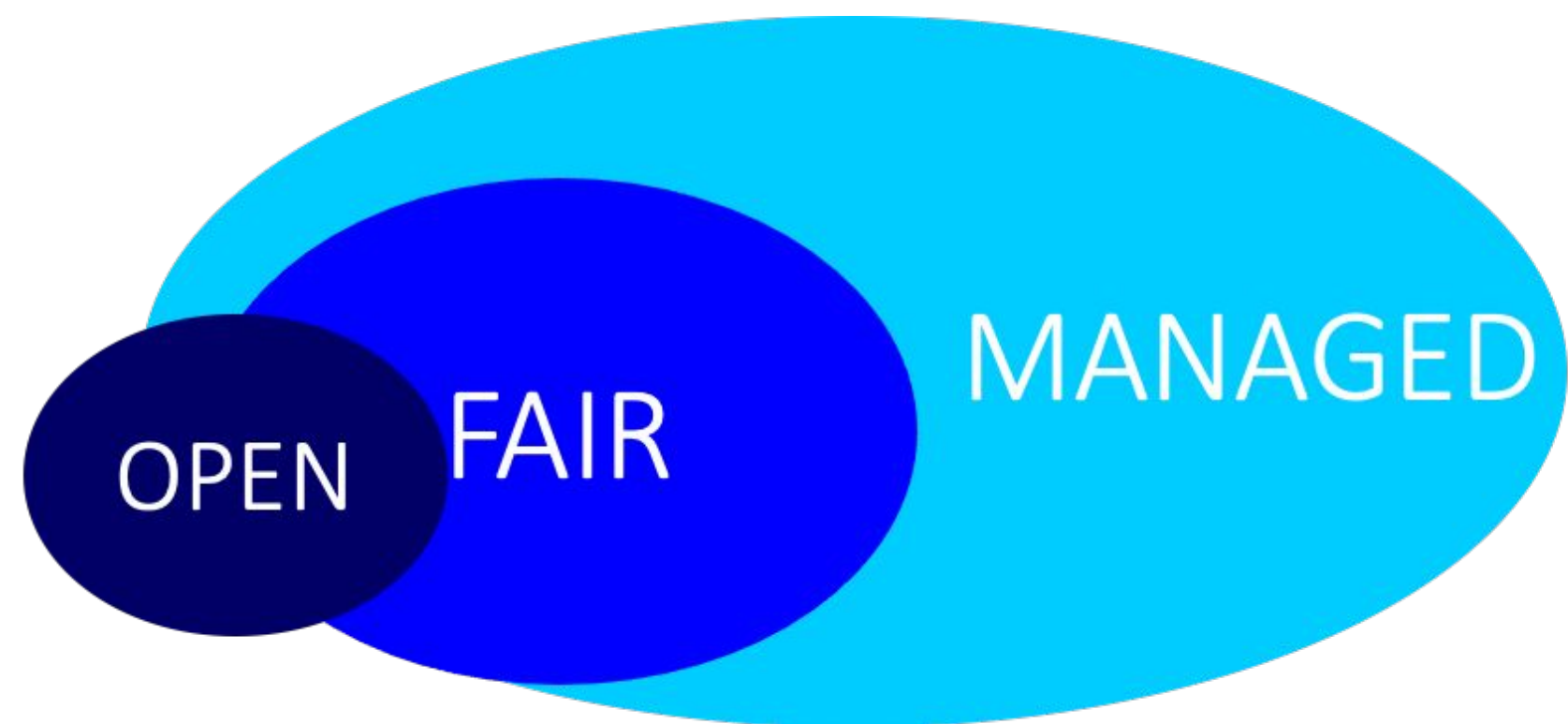
Open science

Fostering open Science to
widen the user community
of scientific data

In the last ten years, OGS has placed the principles of Open Science at the **heart of its action** and in the last five years has stepped up its push towards adopting its principles, with the definition of programs for the **enhancement of Open Science in marine, geophysical, seismological and polar fields.**

The "Open Science" mission is developed following 5 scientific priorities covering the different areas of expertise of the organisation:

- **Findable, accessible, interoperable, reusable (FAIR) data;**
- Risk education and ocean literacy;
- High-performance computing, artificial intelligence and digital twins;
- Citizen Science;
- Science and environmental diplomacy.



ENEA

Consiglio Nazionale
delle Ricerche

Open data are data that anyone can **access, use** and **share**. Possibly subject to the need to cite the source and to share them with the same type of **license** with which they were originally released.

F = Findable: **Metadata** and data should be easy to find for both humans and computers (data are assigned a globally unique and persistent identifier, described with rich metadata, indexed in a searchable resource)

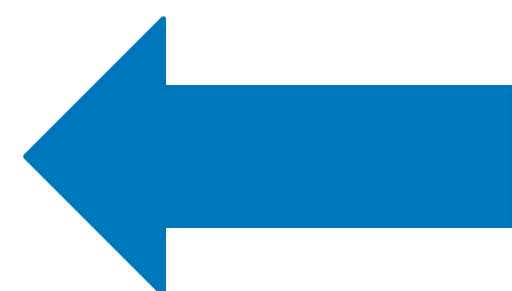
A = Accessible: Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including **authentication** and authorisation.

I = Interoperable: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation and **vocabularies** that follow FAIR principles

R = Reusable: (Meta)data are released with a clear and accessible data **usage license**

Licence

Need to define data access rights or data policy



A Data Policy was defined in 2019 for the Division of Oceanography

- ❖ It needs to be updated, extended to include data from other OGS divisions and integrated in **OGS Data Management Plan (DMP)**



HORIZON 2020 ONLINE MANUAL

“DMPs are a **key element** of good data management...describes the data management life cycle for data that is processed and/or generated by a **Horizon 2020 Europe**).

DMP should include information on:

- the handling of research data during & after the project
- what data will be collected, processed and/or analysed
- which methodology & standards will be applied
- whether data will be shared/made open access
- how **data will be curated & preserved** (including after the project).



M. Lipizer: Open science for societal change

ISTITUTO NAZIONALE
DI OCEANOGRAFIA E DI GEOFISICA SPERIMENTALE



Sgonico, 21 novembre 2019

Documento finale della Commissione Dati di OCE

CRITERI GENERALI DI ARCHIVIAZIONE E GESTIONE DATI, E RELATIVA POLITICA DI ACCESSO

DMP: so many benefits



Make data FAIR

- Makes structuring and documenting of your datasets simpler, thus making it easier for others as well as your future self to find and understand the material;
- Encourages you to think about the data format which is best suited for reuse;
- Allows you to think about the reuse license you would want to apply to your data;
- Choose a proper repository etc.



Clarifies needed budget

- calculating time and resources for careful documentation as well as server space, backup solutions, hardware and software etc.
- Calculating time and resources (money and expertise) for collecting, analysing, and publishing on data.



Allows for easy project management

- An important function of a DMP is to work as a one-stop shop to find project-related information.
- Questions surrounding data management are being gathered in one place and project-related details are readily available rather than just vaguely remembered or simply forgotten.

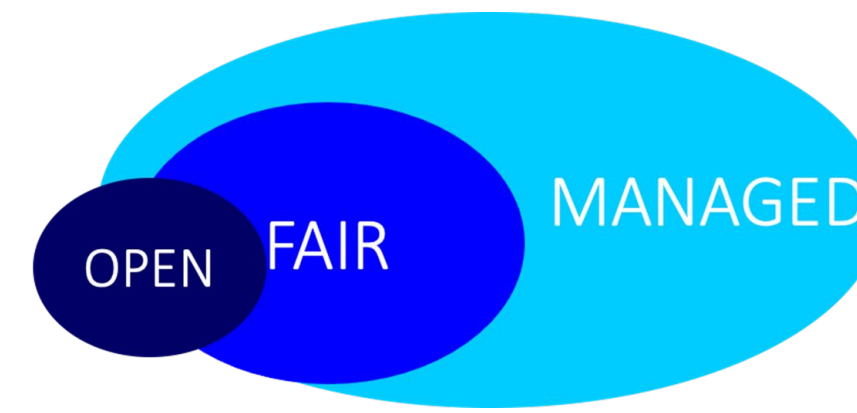


Shows accountability

If you draw up a DMP, you are showing your affiliated institution, funders and project partners a serious approach to research data management, that includes a responsible approach towards research funds and research participants.

<https://dmeq.cessda.eu/Data-Management-Expert-Guide/1.-Plan/Benefits-of-data-managem>

Data curation – data steward



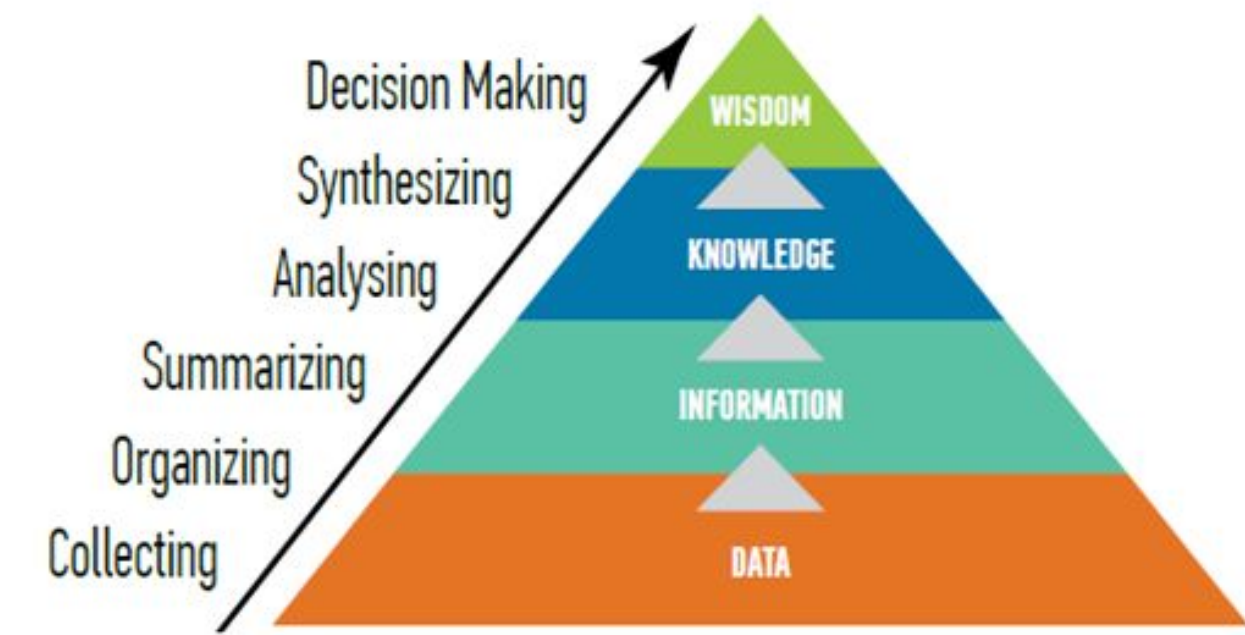
.... Data don't become FAIR by magic!

Data Steward: responsible for the day-to-day activities around the data including the **quality** of the data, ensuring its **safe archival and storage**, and providing the required **metadata** and **documenta...**



Source: [Australian National Data Service](#)

Data curation: from good data ... to good decisions

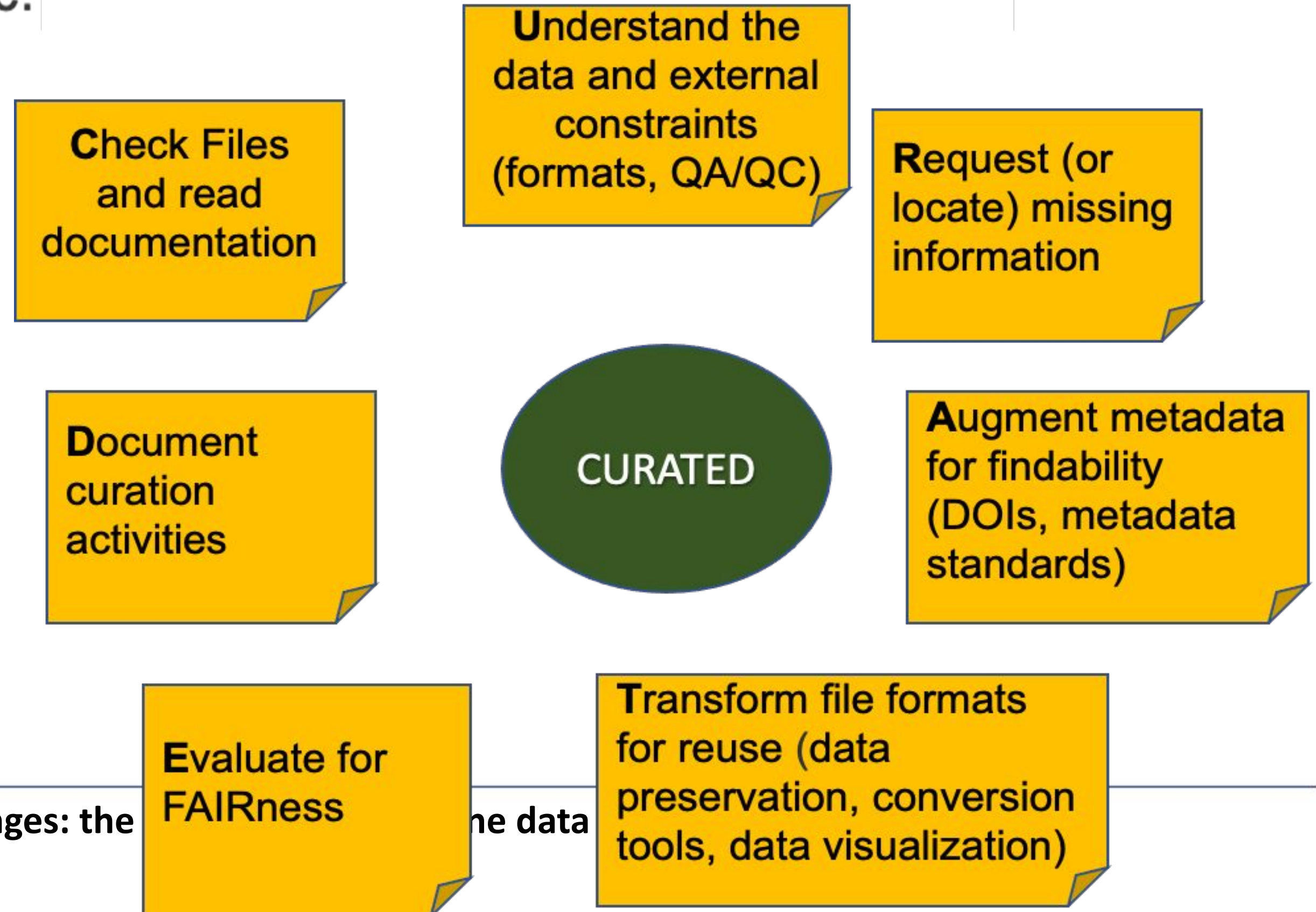


What is data curation?

Data Curation is the documentation, management, and preservation of research data to produce datasets that are FAIR: Findable, Accessible, Interoperable, and Reusable.



cc-by <http://hochsteternbach.wordpress.com>



Data managed:

**Complexity,
heterogeneity**

Oceanographic data:

Physical oceanography

Chemical oceanography (nutrients, dissolved gases, ocean acidity, chemical pollution, marine litter,...)

Biological oceanography

From different monitoring systems:

Oceanographic surveys (cruises,...)

Fixed platforms (oceanographic buoys, eulerian data...)

Free floating (Lagrangian) instruments

Autonomous vehicles

Citizen science

....

With different spatial and temporal resolutions:

Real time data

Delayed-mode data

For different purposes:

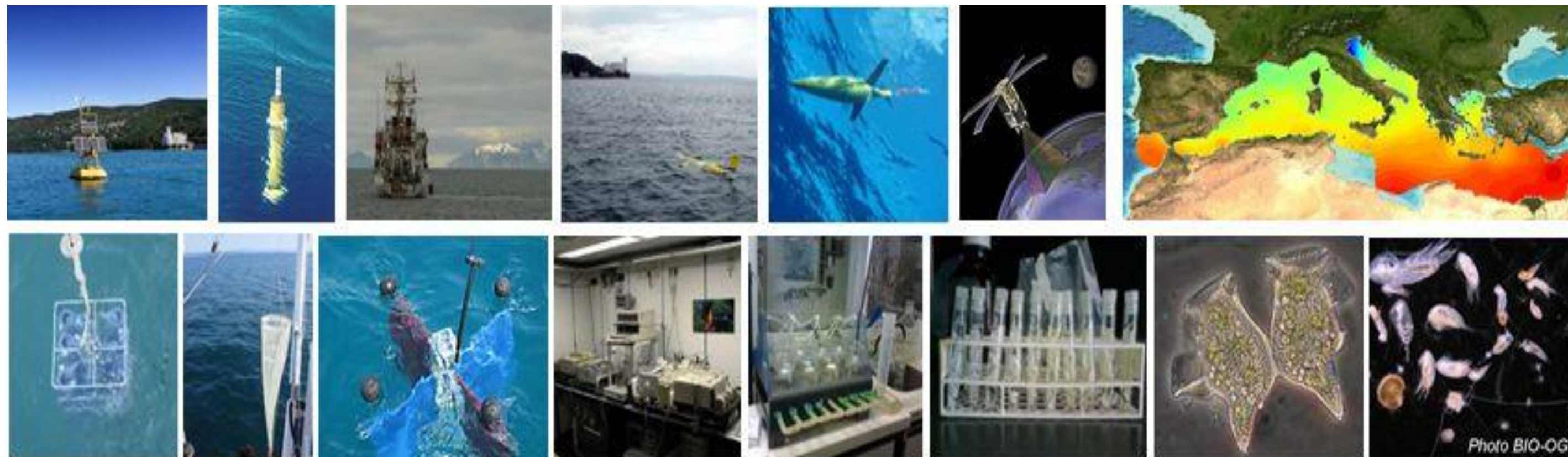
Research data

Operational systems

«Industrial data»

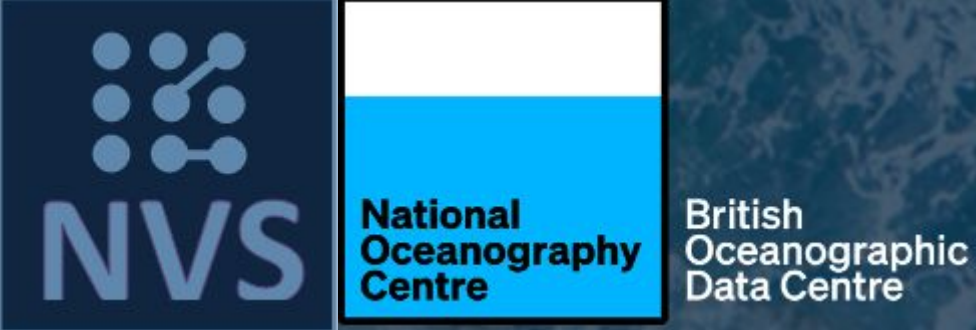
Data management and validation requires multiple expertise and strong collaboration

- ❖ Data curation
- ❖ Data stewards



How data become FAIR:

- ❖ Use of **consolidated standards** (e.g vocabularies, metadata,...) developed within long-term EU programs in collaboration with the European network of National Oceanographic Data Centers
- ❖ Use of persistent identifier (**Digital Object Identifier**) (promotes data citation and originator acknowledgment)
- ❖ Use of **consolidated data access programs**



- ❖ OGS part of a large network of Oceanographic Data Centers



"Environmental Research Division Data Access Program»

ERDDAP

ERDDAP is a data server that gives you a simple, consistent way to download subsets of scientific datasets in common file formats and make graphs and maps. This particular ERDDAP installation has oceanographic data (for example, data from satellites and buoys).



Ongoing work to connect and align metadata with major international data infrastructures: **World Ocean DataBase, PANGEA, Copernicus, MSFD** and **WFD** data collection framework



From good data ... to good decisions: data quality

Data Quality Control (QC) is a crucial step

NODC applies:

Consolidated semi-automatic QC routines for “Real Time” **meteo - oceanographic** data and for “Delayed mode” oceanographic data

Ongoing work on update and revision of:

- ❖ **biogeochemical** QC, update of climatologies and concentration ranges in coastal and open waters
- ❖ Definition of QC protocols to be used at EU level

Biodiversity data: use of EU standards for metadata and taxonomy and standard metadata QC

Need of dedicated biodiversity data curation

Citizen science: an example of active society engagement and science co-creation

Requires dedicated QC procedures



Gulf of Trieste observing system
Regional Civil Protection
Contribution to North Adriatic Digital Twin (PNRR)

From good data ... to good decisions: data quality

Marine pollution: chemical pollution and marine litter

Chemical pollution: limited standard QC procedures

- ❖ Efforts focused on collecting required QA, QC and methodological information to evaluate **data quality & comparability**
- ❖ Definition of concentration ranges of different substances in specific areas
- ❖ Ongoing work to propose QC protocols to be used at EU level
- ❖ **Challenges:** very high and increasing number of pollutants, limited knowledge

Marine litter: limited standard procedures

- ❖ Use of EU standards for metadata and data
- ❖ Automated/Semi-automated semantic and syntactic check procedures for data and metadata, and for «impossible metadata (position, timestamp...)»
- ❖ **Challenges:** some new and non-consolidated monitoring data with not clear needs and requirements, need of QA, QC and methodological information to evaluate data quality & comparability ...lacking of efficient link with contaminants/chemicals/additives



Heavy Metals in the Adriatic-Ionian Seas: A Case Study to Illustrate the Challenges in Data Management When Dealing With Regional Datasets

Maria Eugenia Molina Jack^{1*}, Rigers Bakiu², Ana Castelli³, Branko Čermelj⁴, Maja Fafandel⁵, Christina Georgopoulou⁶, Giordano Giorgi⁷, Athanasia Iona⁸, Damir Ivankovic⁹, Martina Kralj¹, Elena Partescano¹, Alice Rotini⁷, Melita Velikonja¹⁰ and Marina Lipizer¹

¹ Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS), Trieste, Italy, ² Department of Aquaculture and Fisheries, Faculty of Agriculture and Environment, Agricultural University of Tirana, Tirana, Albania, ³ Public Institution of Marine Research, Dubrovnik, Croatia, ⁴ National Institute of Biology, Marine Biology Department, Ljubljana, Slovenia, ⁵ Institute of Biology, Marine Biology Department, Zagreb, Croatia, ⁶ Region



Marine Pollution Bulletin
Volume 185, Part A, December 2022, 114181



Trace metals and polycyclic aromatic hydrocarbons in the Eastern Mediterranean sediments: Concentration ranges as a tool for quality control of large data collections

M. Lipizer^a, D. Berto^b, B. Čermelj^c, M. Fafandjel^d, M. Formalewicz^b, I. Hatzianestis^e, N. Iljanić^f, H. Kaberi^e, M. Kralj^a, S. Matijevec^g, M.E. Molina Jack^a, C. Parinos^e, J. Tronczynski^h, M. Giani^a

Show more

+ Add to Mendeley Share Cite

<https://doi.org/10.1016/j.marpolbul.2022.114181>

Under a Creative Commons license

Get rights and content



JRC TECHNICAL REPORTS

EU Marine Beach Litter Baselines

Analysis of a pan-European 2012-2016 beach litter dataset

Hanke, G., Walvoort, D., Van Loon, W., Addamo, A.M., Brosich, A., del Mar Chaves Montero, M., Molina Jack, M.E., Vinci, M., Giorgetti, A.

MSFD Technical Group on Marine Litter
2019



Open science for societal challenges:

- ❖ Do we have the data we (scientific community, society, policy,...) need?
- ❖ Are they fit – for – purpose?



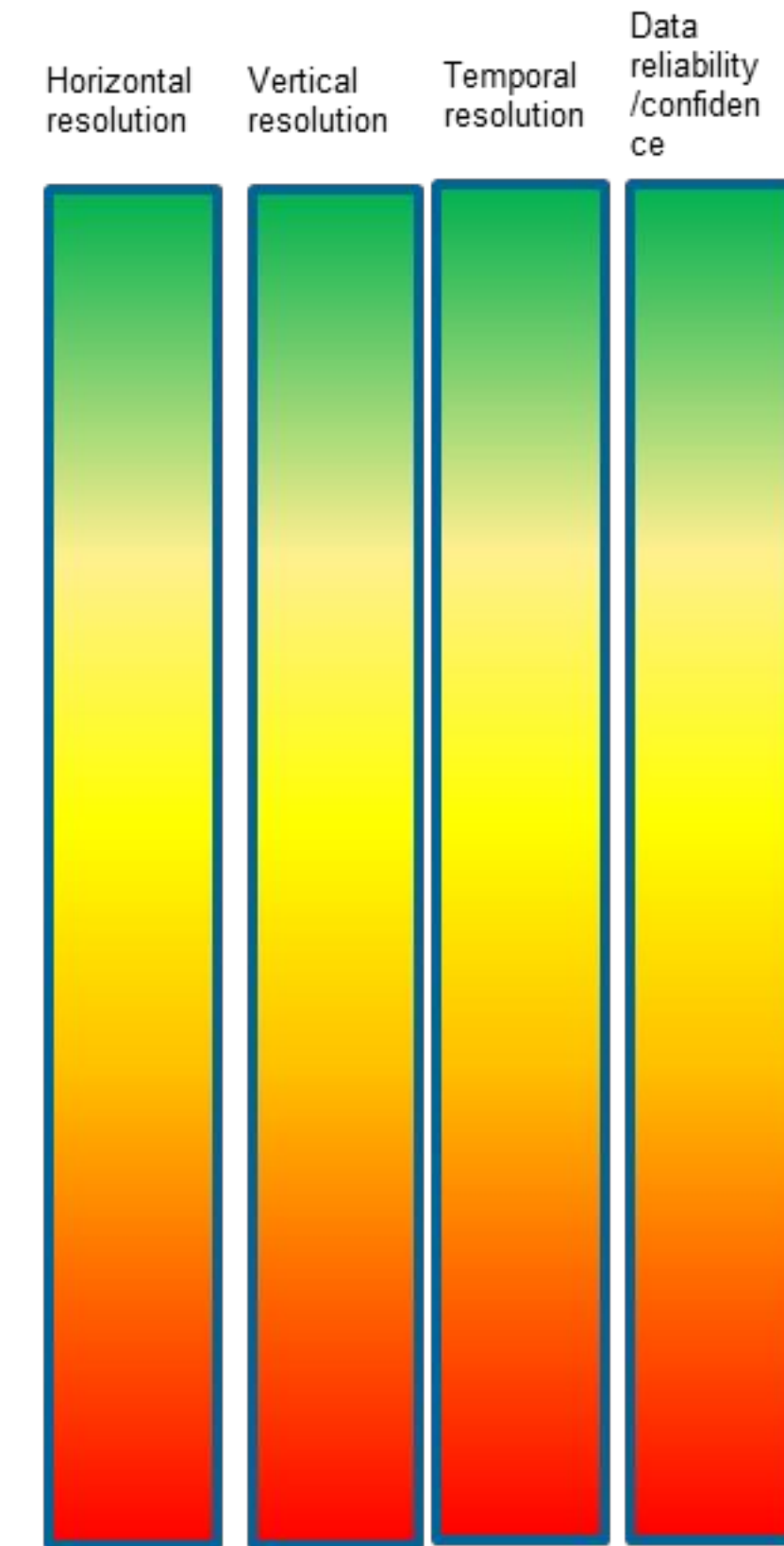
Threats

Observing systems



	Discrete data - <i>in situ</i> sampling	Continuous data - Fix platforms	Autonomous vehicles	Lagrangian platforms	... Omics...	Satellite	Prediction (modelling)	User-friendly information (policy - society)
Warming	[Color gradient: green to red]							
Acidification	[Color gradient: yellow to red]							
Sea level rise	[Color gradient: yellow to red]							
Erosion	[Color gradient: yellow to red]							
Noise	[Color gradient: yellow to red]							
Eutro/oligotrophication	[Color gradient: yellow to red]							
Pollution	[Color gradient: yellow to red]							
Oil spill	[Color gradient: yellow to red]							
Marine litter	[Color gradient: yellow to red]							
Non Indigenous Species/Invasive	[Color gradient: yellow to red]							
Harmful Organisms	[Color gradient: yellow to red]							
Habitat loss	[Color gradient: yellow to red]							
Biodiversity loss	[Color gradient: yellow to red]							

Fitness for purpose



Open?
FAIR?

M. Lipizer: C

Are current systems fit for purpose?

marin

→ Main gaps

Open science for societal challenges:

Stakeholder dialogue within several projects (HarmoNIA, SHAREMED, EMODnet,...), with “policy” communities (e.g. Marine Strategy Framework Directive, Maritime Spatial Planning, EU Biodiversity Strategy, UN – Ocean Decade, EU Zero Pollution Strategy,...)

⇒ To identify “the science we need for the ocean we want”

⇒ Stakeholder needs:

- FAIR data
- “Transparency” in data (quality, methodological information, access conditions, acknowledgment,...)
- user-friendly marine indicators
- user-friendly data products

→ Integrating data and information towards the Digital Twin of the North Adriatic (PNRR iNEST)



2021-2030 United Nations Decade of Ocean Science for Sustainable Development



Open Science

Ongoing development to better fit societal needs:

- ❖ Contribution to the European Open Science Cloud (EOSC)
- ❖ Contribute to implementation of Virtual Research Environment (VRE)
- ❖ Contribution to implementation of Digital Twin of the Ocean



Federated and open multi-disciplinary environment

The European Open Science Cloud

Providing 1.8 million European researchers and 70 million professionals in science and technology a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines.

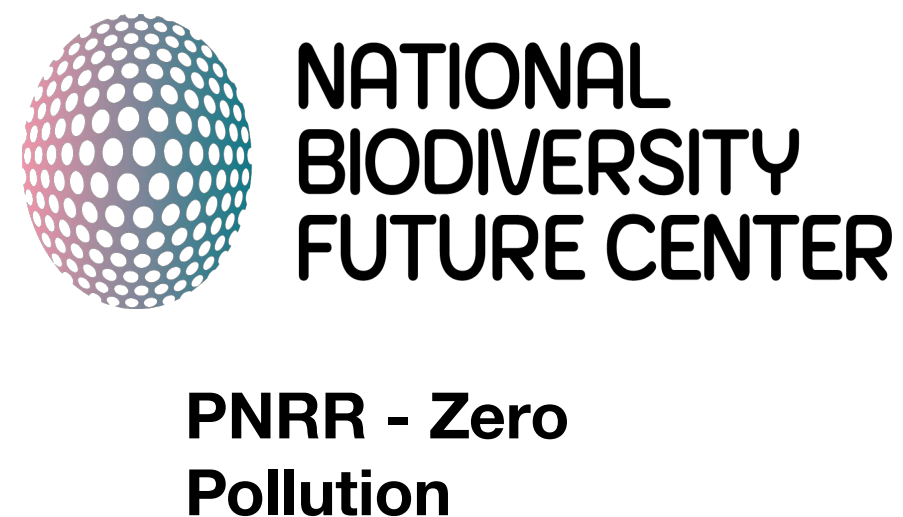


Bottlenecks:

- ❖ We need additional Data Stewards but: difficulties in hiring! Really difficult to find the required IT expertise
- ❖ Still a lot of manual processing due to data in non-standard formats
- ❖ Limited awareness and acknowledgment of the importance (and efforts required!) of marine data curation

Opportunities:

- ❖ **OGS expertise** required by many initiatives: EU Technical Group Marine Litter, UNEP Global Partnership on Marine Litter (GMPL), International Marine Debris Data Harmonisation initiative (Ministry of the Environment, Japan), JRC for MSFD, UNEP/MAP MEDPOL, growing role in EMODnet Biology, Back to Blue Initiative (Nippon Foundation and the Economist Group)
- ❖ The demands of OGS and of several national (e.g. PNRR) and international initiatives (e.g. JRC for MSFD) to manage an **ever-increasing amount** and **variety** of data are constantly growing



HORIZON-CL6-2023-ZERO POLLUTION-01



An initiative of
The Economist Group and The Nippon Foundation
WHY SHOULD WE CARE ABOUT THE OCEAN POLLUTION DATA GAP?

Additional data stewards strongly needed to turn marine Open Science into reality!



M. Lipizer: Open science for societal challenges: the importance of marine data curation

